Georeferencing the GESIS Panel: Background, Workflow, and Analysis Example

Jünger, Stefan; Kolb, Jan-Philipp; Schwerdtfeger, Maikel

gesis Leibniz-Institut für Sozialwissenschaften

GESIS Papers

2020|10

Georeferencing the GESIS Panel: Background, Workflow, and Analysis Example

Stefan Jünger, Jan-Phillip Kolb & Maikel Schwerdtfeger

GESIS Papers 2020|**10**

Georeferencing the GESIS Panel: Background, Workflow, and Analysis Example

Stefan Jünger, Jan-Phillip Kolb &
Maikel Schwerdtfeger

GESIS – Leibniz-Institut für Sozialwissenschaften 2020

GESIS Papers

GESIS – Leibniz-Institut für Sozialwissenschaften
Datenarchiv für Sozialwissenschaften
Data Linking and Data Security
Unter Sachsenhausen 6-8
50667 Köln

E-Mail: stefan.juenger@gesis.org

ISSN: 2364-3781 (Online)
Herausgeber,
Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

1 Introduction

Georeferencing means adding direct spatial identifiers, such as geo-coordinates, to data (Meyer & Bruderer Enzler, 2013). While this technique is mainly associated with disciplines of the geosciences, in recent years, social science survey research also has shown substantial interest in using spatial methods (Bluemke et al., 2017), following the promises of analyzing small-scale geospatial information jointly with existing survey attributes (Hillmert et al., 2017; Schweers et al., 2016). For example, by having access to georeferenced survey data, researchers could locate survey respondents in space to control for spatial clustering, thus improving estimates of analyses, and they can incorporate small-scale geo-related hypotheses in their research questions (Müller et al., 2017). Accordingly, what has been emerging are numerous applications in an extensive range of social science subdisciplines, from ethnicity (Sluiter et al., 2015; Tolsma & van der Meer, 2017), political attitudes (Förster, 2018; Klinger et al., 2017), and health (Boes et al., 2013; Sørensen et al., 2013), to the family (Downey et al., 2016), education (Roos et al., 2013; Weßling, 2016) and inequalities research (Crowder & Downey, 2010; Zwickl et al., 2014).

This paper describes how we handle the process of georeferencing for the survey data of the GESIS Panel. Georeferencing these data is part of a larger initiative at GESIS concerned with spatially integrating existing survey programs (Schweers et al., 2016). In this initiative, subsequently, all data of the national survey programs of GESIS, e.g., the German General Social Survey (GGSS) (Klinger, 2018) or the German Longitudinal Election Study (GESIS - Leibniz Institute for the Social Sciences, 2019) were georeferenced and prepared for spatial linking projects. As part of these survey programs, the GESIS Panel is one of the most comprehensive panel data collections.

We organized the paper as follows. First, we introduce the basic terms that are necessary to describe the general procedure of georeferencing. We then proceed to display specifics of the GESIS Panel and its mixed-mode survey structure. The subsequent section illustrates how the addresses of the GESIS Panel respondents were geocoded, most importantly, what measures of data protection we established, and how the workflow now in place will be continued in the future. Overall, this already exemplifies the georeferencing the GESIS Panel. Therefore, to highlight some of the advantages of using georeferenced survey data in the first place, we conclude this paper with a brief analysis of the GESIS Panel enriched with land use data.

2 Basic Terms, Data, and Concepts

2.1 Georeferencing, Geocoding, and Data Privacy

There are several prerequisites for georeferencing—having access to information about spatial references is one of the essential ones. In the survey data context, these references can be housing addresses, e.g., of survey respondents, or names of municipalities and city districts (Hillmert et al., 2017). Since the purpose of this paper is to exemplify the use of small-scale references from the GESIS Panel, we yet further only concentrate on references based on addresses. In any case, the first step in georeferencing depicts locating observations in data in geographic space, which requires having information about these locations.

Furthermore, we then usually differentiate between indirect spatial references and direct spatial references. The first class refers to spatial references that depict text strings of addresses or municipality names. The second class refers to more structured identifiers of spatial units such as municipality codes, zip codes, or geo-coordinates. Effectively, to conduct spatial analyses, direct spatial identifiers in the form of geo-coordinates are required since they enable researchers to project different data layers into a joined coordinate space and to relate them to each other. In this paper, therefore, we understand the term of georeferencing, as defined in the beginning, as adding geo-coordinates to observations in data.

Consequently, what is needed to apply georeferencing techniques are geo-coordinates. Ordinary survey projects, however, only have access to addresses, which are indirect spatial references. While there are developments to use, e.g., GPS devices to receive geo-coordinates automatically during interviewing respondents (Bluemke et al., 2017), applications of this approach are still rare. Accordingly, addresses have to be converted into geo-coordinates, which is known as geocoding (Zandbergen, 2014). The lack of geo-coordinates in regular surveys yet implies some additional work and demands since survey projects need access to methods of geocoding, which are not trivial to perform.

Several providers offer geocoding as a service. Some of the survey institutes that conduct surveys, such as Infas (https://infas360.de/infas-geodaten/), already geocode addresses. Alternatively, researchers can geocode addresses on their own by using one of the third-party services on the internet. For instance, the geocoding service from Google has often been used widely, but Google has recently changed the terms of use, so that use of the geocoder is now only possible with registration. Moreover, researchers should keep in mind that Google might store requests to their service comprising address information of survey respondents who should be protected. As an alternative, the Nominatim API of the OSM project also can be used to perform geocoding (Warden, 2011). Generally, accessing a geocoding service implies pushing sensitive information (i.e., survey respondents' addresses) to third parties, which is why, before researchers use specific services for geocoding, they should evaluate the consequences of data protection.

At GESIS, we have licensed the geocoding service of the Federal Agency for Cartography and Geodesy (BKG). The BKG geocoder processes the addresses on the fly: it does not store the content of the requests, i.e., the addresses and metadata belonging to the request. Consequently, it prevents us from disclosing any personal information of survey respondents to the BKG and therefore complies with the data protection legislation in Germany. GESIS uses the BKG geocoder for all its geocodings of survey addresses and offers it likewise as a service to other researchers as-is and on their own risk.

Meanwhile, German data protection legislation remains strict. If we aim to use georeferenced information based on personal information, such as addresses, we must pay very close attention to data privacy. Survey attributes (i.e., the answers to survey questions) must not be stored together with personal information (i.e., addresses and their geo-coordinates), and research projects are usually conservative in this manner. They store survey data and address data separately; only tables with strict access rights allow establishing correspondence in some cases, e.g., in a panel data scenario where they aim to add newly collected data to previous study waves. Concerning these arrangements, georeferenced survey data do not differ from ordinary survey data.

What georeferenced survey data differ from, however, are other geospatial data that can be displayed in the form of maps (see soil sealing as an example below). Typically, these data in the form of geospatial data formats store location information and associated attributes in one file. This style of storing is usually not carried out with georeferenced survey data. Instead, if researchers aim to use georeferenced survey data for the spatial linking purpose, they link geo-coordinates to the geospatial data, extract the geospatial information and then add them to the survey data (Schweers et al., 2016). Only the last step requires using a correspondence table after deleting the geo-coordinates from the linked dataset. Georeferenced survey data are an exceptional case of georeferenced data, and researchers have to apply additional steps in using them (Müller, 2019).

To sum up, georeferenced survey data are survey data after geocoding respondents' addresses. These data open new opportunities for methods that go beyond the mere localization of survey respondents in space, as we will demonstrate in section 5. However, dealing with such small-scale geospatial information of survey respondents poses some challenges to preserve data privacy. Georeferenced survey data are consequently not yet ready to use after geocoding addresses and must be treated with caution.

2.2 Geospatial Data and Spatial Linking

Georeferenced survey data not only provide fine-grained information about respondents' locations; they are also used to combine survey data with data from other sources. In order to perform these combinations, spatial linking techniques (Müller et al., 2017) available in Geographic Information Systems (GIS) (Bluemke et al., 2017), are the methods of choice. Spatial linking methods, also called spatial join or overlay, offer a vast range of geospatial operations to combine georeferenced survey data and geospatial data from auxiliary sources.

To get an idea of what kind of operations these techniques involve, a closer look at the properties of geospatial data is helpful. Geospatial data comprise information on geometries, such as points, lines, polygons, or uniformly shaped grid cells. Accordingly, by using spatial linking methods, researchers can gather information about these different geometries. For example, they can analyze if people live in proximity to a point or a line. These geometries, as shown in *Figure 1*, then can comprise additional information: points can depict the location of a kindergarten, a line the course of a noisy road, polygons the extent of a neighborhood with a specific unemployment rate, and a grid cell information on land use. In combination with their location in space, geospatial data provide complex additional information, e.g., by calculating distances to the next kindergarten for each survey respondent.

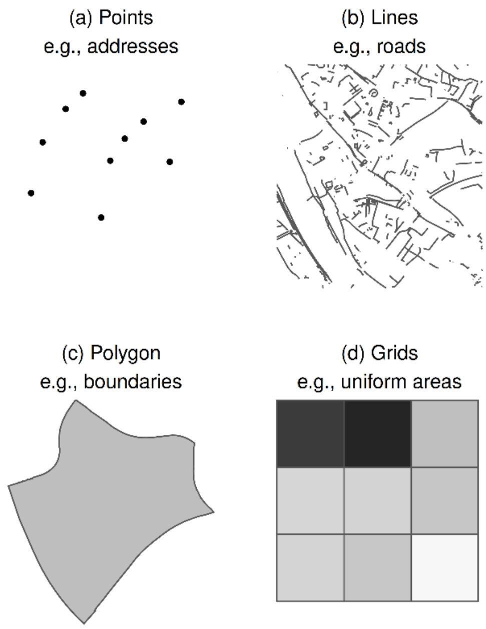


Figure 1: Different types of geometries available in geospatial data (Source: Jünger, 2019)

In this paper, we use data from the Leibniz Institute for Urban and Regional Development (IOER) and its IOER Monitor, which offers longitudinal geospatial data on land use in Germany. These geospatial data comprise of uniformly shaped grid cell information and, therefore, are only one example of geospatial data that social science researchers potentially can use for their spatial linking projects, also with the georeferenced GESIS Panel. For other examples of geospatial data that were spatially linked to the georeferenced GGSS, please refer to Müller et al. (2017) or Jünger (2019).

3 Overview of the GESIS Panel

3.1 Survey Structure

The GESIS Panel is a probability-based mixed-mode access panel that started in 2013 (Bosnjak et al., 2018). What does that mean?

Probability-based means that the selection of persons is based on a random sample from the German population registers. All German-speaking persons residing in private households, who are registered in Germany, and who are between 18 and 70 years belong to the target population of the GESIS Panel. Specifically, a stratified random sampling approach was applied, that is, municipalities were sampled randomly, and proportional-to-size as primary sampling units in the first stage, whereas individuals were sampled in a second stage. Besides, refreshments for the GESIS Panel are realized every two years. These are based on the German General Social Survey (GGSS), which has the same sampling routine. Participants in the GGSS are asked if they are willing to also participate in the GESIS Panel.

The mixed-mode in the description of the GESIS Panel means that there are two groups of participants. The respondents, our panelists, can choose whether they want to participate via an offline questionnaire by mail or online using their web browser. This arrangement ensures that people who cannot or do not want to participate via the Internet are not excluded from the panel.

Lastly, the access panel means that researchers can submit their studies to the GESIS Panel. The proposals are then peer-reviewed, and if the review is positive, researchers can have a five-minute slot per requested panel wave for their study. Thus, the study can be either cross-sectional or longitudinal. To summarize, the GESIS Panel had fielded over 90 studies in 41 waves by the beginning of the year 2020 due to its constant data collection every two months. The data is available as scientific use file ("Standard Edition") and as an "Extended Edition" that is only accessible for on-site use in the GESIS Secure data center (SDC).[1]

3.2 Analysis Potential for Spatial Analysis

Since the topics addressed in the GESIS Panel are various, e.g., ranging from political attitudes to specific work conditions, here we only concentrate on the potential for analyzing questions regarding the spatial surrounding of people. Questions were already asked in the welcome survey of the GESIS Panel concerning knowledge of the direct living environment, which is particularly valuable for environmental research. In this first survey, the panelists are asked, for example, whether they feel affected by environmental influences, such as missing accessible public parks or noise and air pollution in their neighborhood. Most notably, the subjective lack of green spaces is intriguing to compare with the actual distribution of green spaces stemming from the data of the IOER Monitor.

Furthermore, several other studies in the later waves of the GESIS Panel exist that offer exciting opportunities for complementation with geospatial data. One example is the study on "Pro-environmental Behavior in High-Cost Situations" (Neumann & Mehlkop, 2018), which were collected in waves "be" (2014) and "cb" (2015). It combines pro-environmental behavior, like buying food at regional farmer markets or choosing ecological alternatives in investment situations, with mon-

[1] https://www.gesis.org/en/services/data-analysis/more-data-to-analyze/secure-data-center-sdc

etary and social costs and benefits. There may be apparent assumptions about how environmental influences affect pro-environmental behaviors, which could be analyzed with such combined data. For example, the behavior of people in rural areas may be different from the behavior of people from urban ones, simply because of differential public transport infrastructure. Another study, "Within-year Dynamics and Cycles in Subjective Well-being", surveyed cognitive and subjective measures of well-being in 13 consecutive waves and includes questions about the satisfaction with respondents' dwellings. The availability of green spaces in the neighborhood might be a factor influencing the overall satisfaction with the dwelling. Overall, the GESIS Panel provides a rich set of information on individual attitudes, behaviors, values, and subjective perceptions of surroundings, which can be supplemented with auxiliary indicators from the spatial sciences.

4 Georeferencing of the GESIS Panel

This section describes the in-place procedure of georeferencing the GESIS Panel. We first demonstrate the geocoding of respondents' addresses and how we apply measures of data protection. Since georeferencing is an ongoing procedure due to the continually growing GESIS Panel, we also present the workflow of adding new geocoordinates to the data as time proceeds, and new study waves are collected.

4.1 Geocoding of Addresses and Data Protection Measures

Georeferenced survey data are survey data where survey respondents' addresses were converted into a geocoordinate via geocoding. The service of choice for geocoding is the BKG geocoder, not only due to its precision and underlying administrative data but also due to its privacy-friendly implementation. All requests to this service through encrypted SSL connections are processed on the fly, which means no addresses are stored on the servers of the BKG. Choosing this geocoding service ensures compliance with German data protection legislation.

In the GESIS Panel, we use the same data protection measures as in the other survey programs of GESIS. For example, it is not allowed to link individual data from surveys with other individual data. During the process of geocoding and spatial linking, none of the personal information (i.e., geo-coordinates) and survey data attributes are compounded. Instead, we apply a multiple-step approach that divides the steps into single one-purpose functions. For example, one function only deals with requesting geo-coordinates for addresses at the geocoding service; another function only extracts spatial information from geospatial data for a set of geo-coordinates. After each of them is applied, it is evaluated whether information can be deleted or coarsened. Accordingly, it is ensured that at no time more than necessary information are stored.

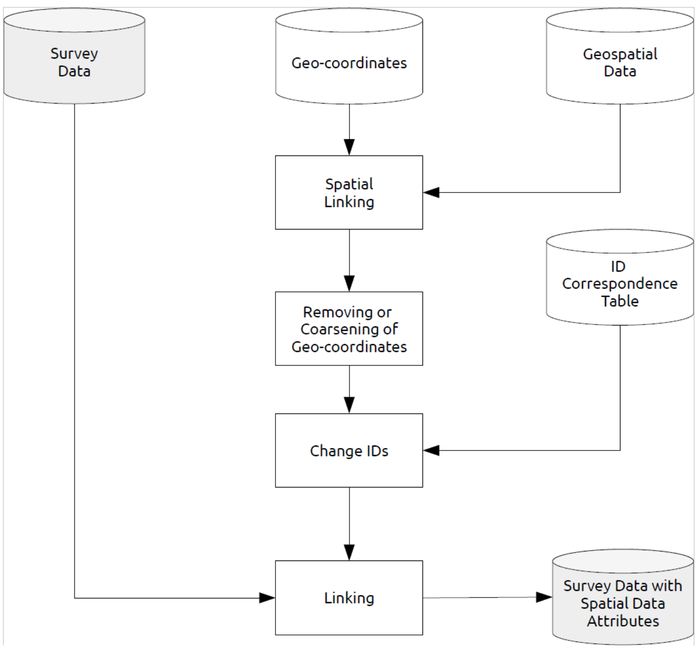


Figure 2: General Concept of Geocoding, Storing and Spatial Linking of Georeferenced Survey Data at GESIS (Source: Jünger, 2019)

Figure 2 displays all details of this process. Broadly, it is divided into three separate process chains, in which the individual functions are applied. The first chain represents the process of geocoding address data using the geocoding service; the second the spatial linking through requesting spatial information (features) from geospatial data sources; and the third the linking of the retrieved information with survey attributes using a correspondence table. A more general description of this process can be found in (Jünger, 2019; Müller, 2019; Schweers et al., 2016).

4.2 Implementation Workflow of Georeferencing

Every two months, a new data collection for the GESIS Panel is in the field. These data collections yield opportunities to gain new information from previous panel respondents, the panelists, and from the ones' that entered the panel after each refreshment. With new respondents also new addresses emerge that require geocoding. Besides other tasks associated with a new data collection, such as documentation and publication, the task of geocoding must also be planned and processed.

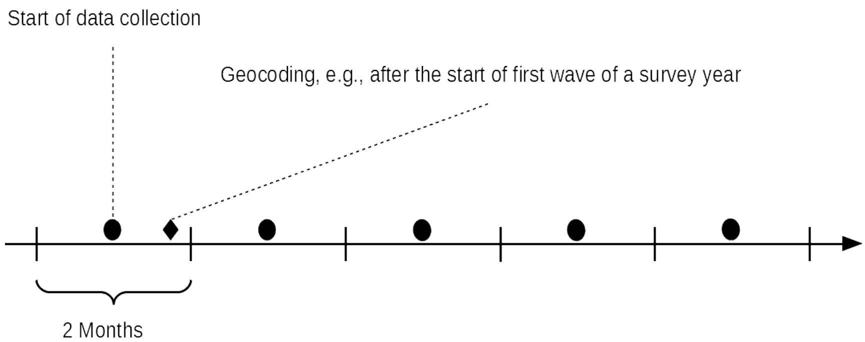


Figure 3: Frequency of Data Collection in the GESIS Panel and Subsequent Geocoding of New Panelists' Addresses

Geocoding at GESIS using the BKG geocoder is now a standardized process. Our staff can rely on an internal R-package's function, which accesses the BKG Application Interface (API) fed by a data frame with addresses and returns a tidy data frame comprising the corresponding geocoordinates. To reduce the strains of all the other tasks, geocoding is only conducted once during the year. *Figure 3* displays this general procedure by drawing a timeline for a typical GESIS Panel data collection year. Each interval exhibits the time of two months in which the data are collected (circular icon) and then once geocoded during the year (diamond icon).

4.3 Data Distribution

In a panel data context, data distribution considering data privacy is a challenging task due to steadily growing information about survey respondents. This fact is somewhat aggravated when such information comprises small-scale geospatial data. Distribution and analysis of the data can only be conducted under strict organizational and technical circumstances. GESIS has established an infrastructure for data containing sensitive information: the GESIS SDC. Data requests and accompanied input data of users are reviewed, their on-site visit for analysis is monitored, and most importantly, which output they aim to take with them, e.g., for writing a paper is controlled by GESIS staff.

Indeed, particularly in times of COVID-19, limited capacities of visits challenge the vision of providing scalable services of data access. According to the FAIR principles, data should be findable, accessible, interoperable, and reusable. To navigate these challenges, the project "Social Spatial Research Data Infrastructure" (SoRa), funded by the German Research Foundation, has laid out some vital groundwork for remote access and remote execution of georeferenced survey data (Bensmann et al., 2020). In this infrastructure, all steps of spatial linking are initiated by users before their actual visit to analyze the data. This procedure limits on-site access time dramatically and does not require to give users access to sensitive information for linking at all.

5 Analysis Example: Spatial Linking with Land Use Data

Georeferenced survey data allows researchers the linking of geospatial information from auxiliary data sources to survey data, which we demonstrate in our analysis example in this section. Both data sources are projected into one coordinate space, while one focal dataset (i.e., the georeferenced survey data) receives information from the other one (i.e., geospatial data) according to the geographic location using Geographic Information Systems (GIS). Researchers can retrieve additional variables for their survey data based on 1:1 locations, proximities, or other measures of spatial analysis. Spatial linking of survey data with other geospatial data sources, therefore, is a flexible method to gain new information for survey respondents.

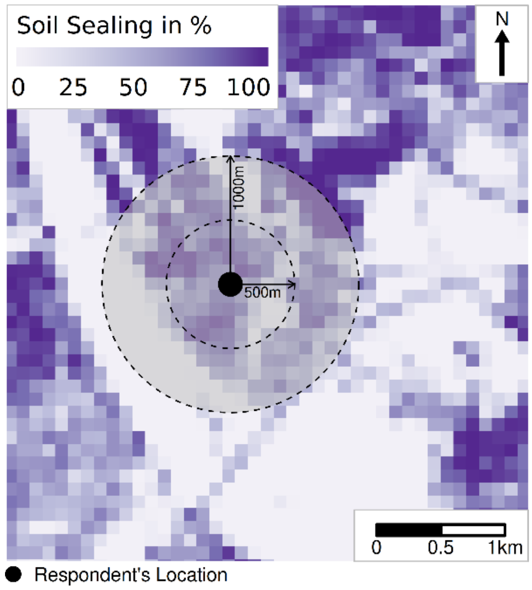


Figure 4: Buffer Calculation of Different Sizes of Soil Sealing for a Single Geo-Coordinate (Source: Jünger, 2019)

Here we only present spatial linking of the georeferenced GESIS Panel with land-use indicators from the Monitor of Settlement and Open Space Development (IOER Monitor) by exploiting spatial buffers. Spatial buffers draw circular areas around a geo-coordinate and support the extraction of descriptive statistics for these areas, such as arithmetic means, maxima, or standard deviations. The spatial buffers can be varied so that researchers also can add geospatial information for different geographic scales to georeferenced survey data. *Figure 4* shows an example of creating spatial buffers around a geo-coordinates of the size 500 and 1000 meters. The information of the underlying data layer, in this case, a soil sealing layer in a 1-hectare grid resolution (see description below), can be added to the geo-coordinate using, e.g., mean values. Spatial buffers are, therefore, a flexible method to model varying environmental influences of people's spatial surroundings.

For our analysis example below, we deploy sizes of 500 meters of spatial buffers to spatially link soil sealing information of the year 2012 (Leibniz Institute of Ecological Urban and Regional Development, 2018) with their mean values to the GESIS Panel respondents of the year 2014 (GESIS - Leibniz Institute for the Social Sciences, 2017). Soil Sealing in percent is the air and watertight coverage of land through roads and buildings. The denser an area is, the higher is the amount of soil sealing. Accordingly, soil sealing is an inverse indicator for green spaces as, by definition, it also involves less free amounts of parks, playgrounds, or other recreational areas.

Table 1: Descriptive Statistics for Soil Sealing in percent after Spatial Linking with the georeferenced GESIS Panel

	Mean	SD
Soil Sealing		
100m Buffer	54.008	24.427
500m Buffer	39.368	23.054
1000m Buffer	30.809	21.875
2000m Buffer	23.204	19.485

Data: Georeferenced GESIS Panel 2014 (GESIS - Leibniz Institute for the Social Sciences, 2017) and IOER Monitor (IOER, 2017)

Table 1 displays some descriptive statistics of soil sealing; however, across different buffer sizes after they were added to the georeferenced GESIS Panel (minimum and maximum values are excluded due to data protection). Generally, soil sealing's indicator values decrease with increasing buffer sizes. Researchers must find measures on how they can evaluate what size of neighborhoods (i.e., buffers) are the most appropriate for their application.

In practice, a variety of approaches are used to test for the most appropriate geographic scale. For example, Sluiter (2015) explicitly proposed the search for the most relevant geographic scale as a research question and subsequently chose the scales with the best fit to the data. While other authors warned to use the statistical model fits as a sole criterium to choose from (Spielman & Yoo, 2009), it seems as least plausible to choose scales that deliver transparent and interpretable pictures about theoretical relationships. In the following, we show an analysis example of the exemplified soil sealing data on how they enrich research with the GESIS Panel.

5.1 Analysis Example

So how can soil sealing data be used in social science research? The idea behind all spatial linking efforts with georeferenced survey data, in general, is that we aim to enrich survey data with additional and meaningful information. Spatial relationships are chosen that are hypothesized to be influential on people. Accordingly, it is always helpful to have measures in the survey data that corroborate this hypothesis by giving some validity to the chosen spatial information. As mentioned, soil sealing comprises an inverse measure of green spaces in neighborhoods, and the GESIS Panel includes questions exactly about annoyance due to the lack of green spaces in respondents' neighborhoods. The relationship between this measure and soil sealing is one example we show in the following.

Another example depicts a more substantive one, which shows the actual impact of soil sealing on the satisfaction with the residence respondents live in. It shows that soil sealing is not only perceived by respondents but also that it has an actual impact. The underlying hypothesis is that green spaces are essential for satisfaction with residences and that the lack of green space reduces them (Bonaiuto & Fornara, 2004).

For both examples, we estimated two separate logistic regression models. In the first model, we used a dichotomized measure of annoyance with the lack of green spaces that was coded with one if respondents answered "strong" or "very strong" on the question "How strongly do you feel affected by the following environmental influences in your residential area? - By lack of accessible green spaces", and zero if they answered "not at all", "low" or "just bearable". In the second model, we used a dichotomized measure of satisfaction with the residence that was coded one if respondents answered "very satisfied" or "rather satisfied" on the question "How satisfied are you - all in all - with your life in [place of residence] at the moment?", and zero if they answered "rather dissatisfied" or "very dissatisfied". As additional survey control variables, we used the age of the respondents in years, their gender (1 = female), and income in 17 categories (with categories ranging from "under 300 Euro" and "300 up to 500 Euro" to "10 000 Euro and more"). The central predictor was the amount of soil sealing in percent as the mean value in 500-meter buffer areas around the respondents. *Table 2* shows the regression table for both models.

Table 2: Logistic Regressions of Soil Sealing on Satisfaction with Residence and Annoyance with the Lack of Green Spaces

	Satisfaction	Annoyance
Sealing of Soils, 500m Buffer	-0.009***	0.033***
	(-0.015, -0.003)	(0.029, 0.038)
Age	-0.004	-0.030***
	(-0.013, 0.006)	(-0.038, -0.023)
Gender (Female)	0.278**	-0.213**
	(0.015, 0.540)	(-0.410, -0.016)
Income	0.102***	-0.088***
	(0.059, 0.145)	(-0.122, -0.053)
Constant	1.498***	0.035
	(0.745, 2.250)	(-0.539, 0.609)
N	2718	2497
Log Likelihood	-825.902	-1239.458

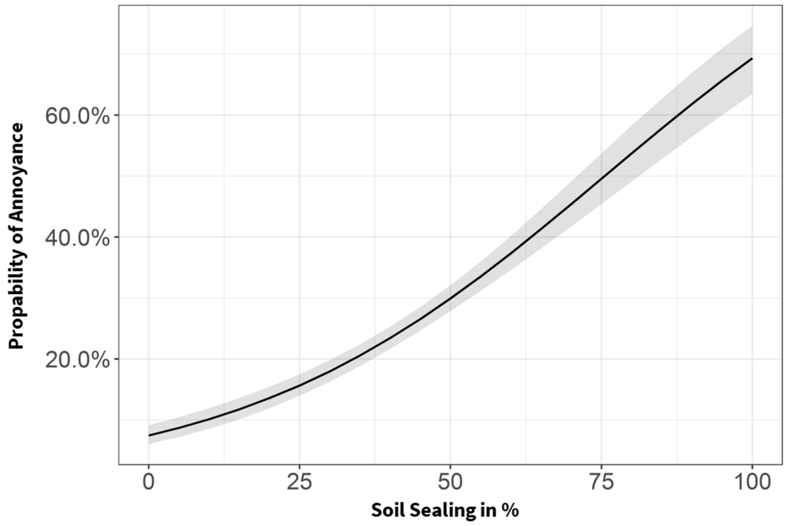
***p < .01; **p < .05; *p < .1

Data: Georeferenced GESIS Panel 2014 (GESIS - Leibniz Institute for the Social Sciences, 2017) and IOER Monitor (IOER, 2017)

Generally, we can see that soil sealing has a negative impact in both models. Soil sealing increases the likelihood of being annoyed with the lack of green spaces, and it decreases the likelihood of not being satisfied with the residence. In contrast, the effect of annoyance is ~3 times higher than for satisfaction. Moreover, the effects of the control variables differ between the models as well. Female and older respondents are less annoyed by the lack of green spaces, but only female respondents are more satisfied with their residence. At the same time, income increases the likelihood of being satisfied with the residence, and it decreases the likelihood of being annoyed with the lack of green spaces. Besides the differential effects of the soil sealing, the control variables also showed some substantial effects on the dependent variables.

We could ask whether these effects for sealing of soils are strong or rather weak as their effect sizes are comparatively small (-.009 for satisfaction with residence and -.033 for annoyance, respectively). Evaluating such a question from a logistic regression model, however, is difficult. Therefore, *Figure 5* and *Figure 6* show the results of computing the predicted probabilities of scoring one on each dependent variable with varying values of soil sealing and fixed values at their mean for all other control variables. In contrast to the effect sizes, these figures show that in both models, soil sealing has a comparatively strong influence on the dependent variables.

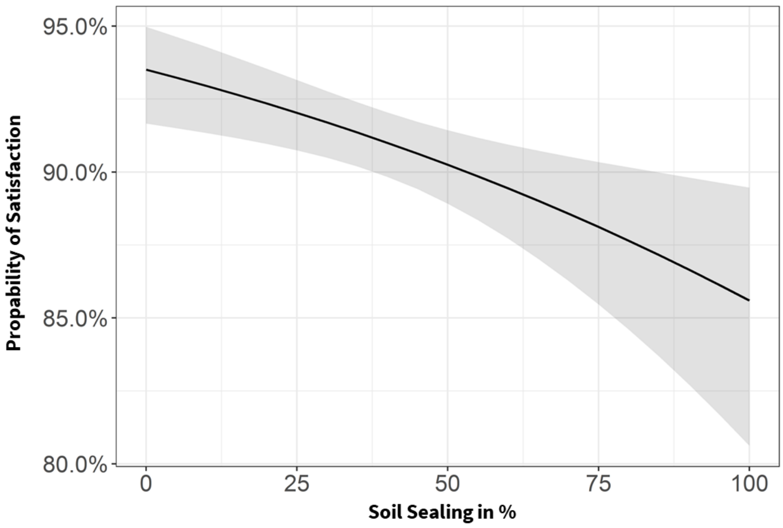
People that live in neighborhoods with low levels of the sealing of soils have a much lower probability of not being annoyed with the lack of green spaces than people that live in neighborhoods with high levels. For example, people that live in neighborhoods with moderate levels of soil sealing of 20% have a probability of 13.76% of being annoyed. In comparison, people that live in neighborhoods with high levels of 80% have a probability of 54.11%. This difference is also visible by the steep increase of probabilities in *Figure 5*.



Data: Georeferenced GESIS Panel 2014 (GESIS - Leibniz Institute for the Social Sciences, 2017) and IOER Monitor (IOER, 2017)

Figure 5: Predicted Values for Annoyance with Lack of Green Spaces Depending on the Amount of Sealing of Soils in 500m Buffer Neighborhoods

Also, the probability of being satisfied with the residence decreases if soil sealing increases. People who live in neighborhoods with moderate levels of soil sealing, e.g., 20%, have a probability of satisfaction of 92.36% that drops to 87.66% in neighborhoods with 80% of soil sealing. The smaller coefficient also reflects this smaller difference in the regression. However, it still shows that soil sealing has a direct influence on dimensions of life satisfaction.



Data: Georeferenced GESIS Panel 2014 (GESIS - Leibniz Institute for the Social Sciences, 2017) and IOER Monitor (IOER, 2017)

Figure 6: Predicted Values for Satisfaction with Residence Depending on the Amount of Sealing of Soils in 500m Buffer Neighborhoods

We also argue that despite these rather small effects, they are substantially significant. Satisfaction with a residence is one dimension of life satisfaction that is rather broad. It involves properties of peoples' apartments, buildings, neighborhoods, and its infrastructure. Lacking green spaces that are associated with soil sealing are only one part of the evaluation process of peoples' residence. Accordingly, a 4.70% difference depicts a considerable drop in the likelihood of being satisfied with the residence due to a difference in soil sealing between 20% and 80%.

6 Conclusion

Combining survey data with spatial information is advertised to increase the knowledge about social behavior and attitudes. Consequently, more and more researchers indeed express their interest in spatially integrating social science theories and data, and research infrastructures and survey projects do well to enable research in this area, if possible. This working paper presented how the GESIS Panel, as one of the largest survey infrastructures in Germany, navigates associated challenges of data protection and distribution for secondary users. Our short analysis example demonstrated that the GESIS Panel offers many interesting studies that would profit from spatially linking small-scale geoinformation to the survey data.

Georeferencing is an enduring process. Due to the panel structure of the GESIS Panel, new respondents regularly enter the sample while others leave the survey. As such, the GESIS Panel must keep track of locational changes of their respondents and must geocode addresses again when necessary. Moreover, the geospatial data landscape in Germany is subject to constant change, and new technological developments (e.g., a change of APIs or geospatial data models) require responses that must be considered. Fortunately, GESIS is well-positioned since tasks such as geocoding are mostly automated or are solved together with other research infrastructures in joined projects, such as the SoRa project.

7 References

Bensmann, F., Heling, L., Jünger, S., Mucha, L., Acosta, M., Goebel, J., Meinel, G., Sikder, S., Sure-Vetter, Y., & Zapilko, B. (2020). An Infrastructure for Spatial Linking of Survey Data. *Data Science Journal*, *19*, 27. https://doi.org/10.5334/dsj-2020-027

Bluemke, M., Resch, B., Lechner, C., Westerholt, R., & Kolb, J.-P. (2017). Integrating Geographic Information into Survey Research: Current Applications, Challenges and Future Avenues. *Survey Research Methods*, *11*(3), 307--327. https://doi.org/10.18148/srm/2017.v11i3.6733

Boes, S., Nüesch, S., & Stillman, S. (2013). Aircraft Noise, Health, and Residential Sorting: Evidence from Two Quasi-Experiments: Aircraft Noise and Health. *Health Economics*, *22*(9), 1037–1051. https://doi.org/10.1002/hec.2948

Bonaiuto, M., & Fornara, F. (2004). Residential satisfaction and perceived urban quality. *Encyclopedia of Applied Psychology*, *3*, 267–272.

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. *Social Science Computer Review*, *36*(1), 103--115. https://doi.org/10.1177/0894439317697949

Crowder, K., & Downey, L. (2010). Inter-Neighborhood Migration, Race, and Environmental Hazards: Modeling Micro-Level Processes of Environmental Inequality. *American Journal of Sociology*, *115*(4), 1110–1149.

Downey, L., Crowder, K., & Kemp, R. J. (2016). Family Structure, Residential Mobility, and Environmental Inequality: Family Structure and Environmental Inequality. *Journal of Marriage and Family*, *79*(2), 535--555. https://doi.org/10.1111/jomf.12355

Förster, A. (2018). Ethnic Heterogeneity and Electoral Turnout: Evidence from Linking Neighbourhood Data with Individual Voter Data. *Electoral Studies*, *53*, 57--65. https://doi.org/10.1016/j.electstud.2018.03.002

GESIS - Leibniz Institute for the Social Sciences. (2017). *GESIS Panel—Extended Edition* [Data set]. GESIS Data Archive. https://doi.org/10.4232/1.12742

GESIS - Leibniz Institute for the Social Sciences. (2019). *GLES Sensitive Regional Data* [Data set]. GESIS Data Archive. https://doi.org/10.4232/1.13263

Hillmert, S., Hartung, A., & Weßling, K. (2017). Dealing with Space and Place in Standard Survey Data. *Survey Research Methods. Special Issue: Uses of Geographic Information Systems Tools in Survey Data Collection and Analysis*, *11*(3), 267--287. https://doi.org/10.18148/srm/2017.v11i3.6729

IOER. (2017). Monitor of Settlement and Open Space Development. *Leibniz Institute of Ecological Urban and Regional Development*.

Jünger, S. (2019). *Using Georeferenced Data in Social Science Survey Research. The Method of Spatial Linking and Its Application with the German General Social Survey and the GESIS Panel*. GESIS - Leibniz-Institut für Sozialwissenschaften. 10.21241/ssoar.63688

Klinger, J. (2018). *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften—ALLBUS Sensitive Regionaldaten* (No. 2018, 15; VARIABLE Reports). GESIS Data Archive. https://doi.org/10.4232/1.13010

Klinger, J., Müller, S., & Schaeffer, M. (2017). Der Halo-Effekt in einheimisch-homogenen Nachbarschaften: Steigert die ethnische Diversität angrenzender Nachbarschaften die Xenophobie? *Zeitschrift Für Soziologie*, *46*(6), 402–419. https://doi.org/10.1515/zfsoz-2017-1022

Leibniz Institute of Ecological Urban and Regional Development. (2018). *Soil Sealing. Monitor of Settlement and Open Space Development.* Retrieved October 2, 2018, from http://monitor.ioer.de/cgi-bin/wcs?MAP=S40RG_wcs

Meyer, R., & Bruderer Enzler, H. (2013). Geographic Information System (GIS) and Its Application in the Social Sciences Using the Example of the Swiss Environmental Survey. *Methoden, Daten, Analysen (mda)*, *7*(3), 317–346. https://doi.org/10.12758/mda.2013.016

Müller, S. (2019). Räumliche Verknüpfung georeferenzierter Umfragedaten mit Geodaten: Chancen, Herausforderungen und praktische Empfehlungen. In U. Jensen, S. Netscher, & K. Weller (Eds.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (pp. 211–229). Verlag Barbara Budrich.

Müller, S., Schweers, S., & Siegers, P. (2017). *Geocoding and Spatial Linking of Survey Data—An Introduction for Social Scientists* (No. 2017, 15; GESIS Paper, pp. 1--29). GESIS -- Leibniz Institute for the Social Sciences. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-52316-9

Neumann, R., & Mehlkop, G. (2018). Umweltentscheidungen als Wechselspiel von Einstellungen, Handlungskosten und situativer Rahmung–ein empirischer Theorienvergleich mit Daten des GESIS Panels. *Zeitschrift Für Soziologie*, *47*(2), 101–118.

Roos, L. L., Hiebert, B., Manivong, P., Edgerton, J., Walld, R., MacWilliam, L., & de Rocquigny, J. (2013). What is Most Important: Social Factors, Health Selection, and Adolescent Educational Achievement. *Social Indicators Research*, *110*(1), 385–414. https://doi.org/10.1007/s11205-011-9936-0

Schweers, S., Kinder-Kurlanda, K., Müller, S., & Siegers, P. (2016). Conceptualizing a Spatial Data Infrastructure for the Social Sciences: An Example from Germany. *Journal of Map & Geography Libraries*, *12*(1), 100--126. https://doi.org/10.1080/15420353.2015.1100152

Sluiter, R., Tolsma, J., & Scheepers, P. (2015). At Which Geographic Scale Does Ethnic Diversity Affect Intra-Neighborhood Social Capital? *Social Science Research*, *54*, 80--95. https://doi.org/10.1016/j.ssresearch.2015.06.015

Sørensen, M., Andersen, Z. J., Nordsborg, R. B., Becker, T., Tjønneland, A., Overvad, K., & Raaschou-Nielsen, O. (2013). Long-Term Exposure to Road Traffic Noise and Incident Diabetes: A Cohort Study. *Environmental Health Perspectives*, *121*(2), 217–222. https://doi.org/10.1289/ehp.1205503

Spielman, S. E., & Yoo, E. (2009). The Spatial Dimensions of Neighborhood Effects. *Social Science & Medicine*, *68*(6), 1098–1105. https://doi.org/10.1016/j.socscimed.2008.12.048

Tolsma, J., & van der Meer, T. W. G. (2017). Losing Wallets, Retaining Trust? The Relationship Between Ethnic Heterogeneity and Trusting Coethnic and Non-coethnic Neighbours and Non-neighbours to Return a Lost Wallet. *Social Indicators Research*, *131*(2), 631–658. https://doi.org/10.1007/s11205-016-1264-y

Warden, P. (2011). *Data source handbook* (1st ed). O'Reilly.

Weßling, K. (2016). *The Influence of Socio-Spatial Contexts on Transitions from School to Vocational and Academic Training in Germany* [Eberhard Karls Universität]. https://doi.org/10.15496/publikation-15222

Zandbergen, P. A. (2014). Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data. *Advances in Medicine*, *2014*, 1--14. https://doi.org/10.1155/2014/567049

Zwickl, K., Ash, M., & Boyce, J. K. (2014). Regional Variation in Environmental Inequality: Industrial Air Toxics Exposure in U.S. Cities. *Ecological Economics*, *107*, 494–509. https://doi.org/10.1016/j.ecolecon.2014.09.013